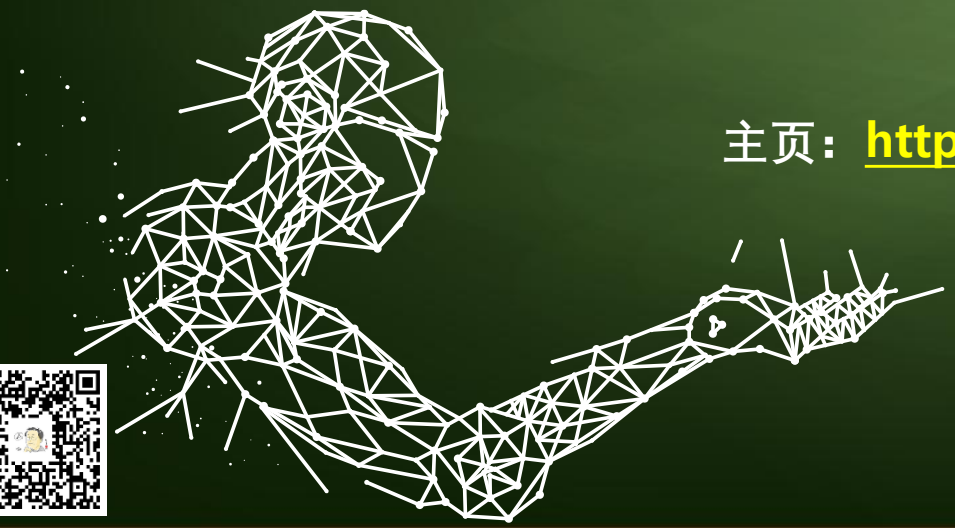
02

第2章 Spark的设计与运行原理

授课教师：[李阳源]
ywlee1212@gmail.com
主页：https://sites.google.com/view/dalguri





第2章 Spark的设计与运行原理



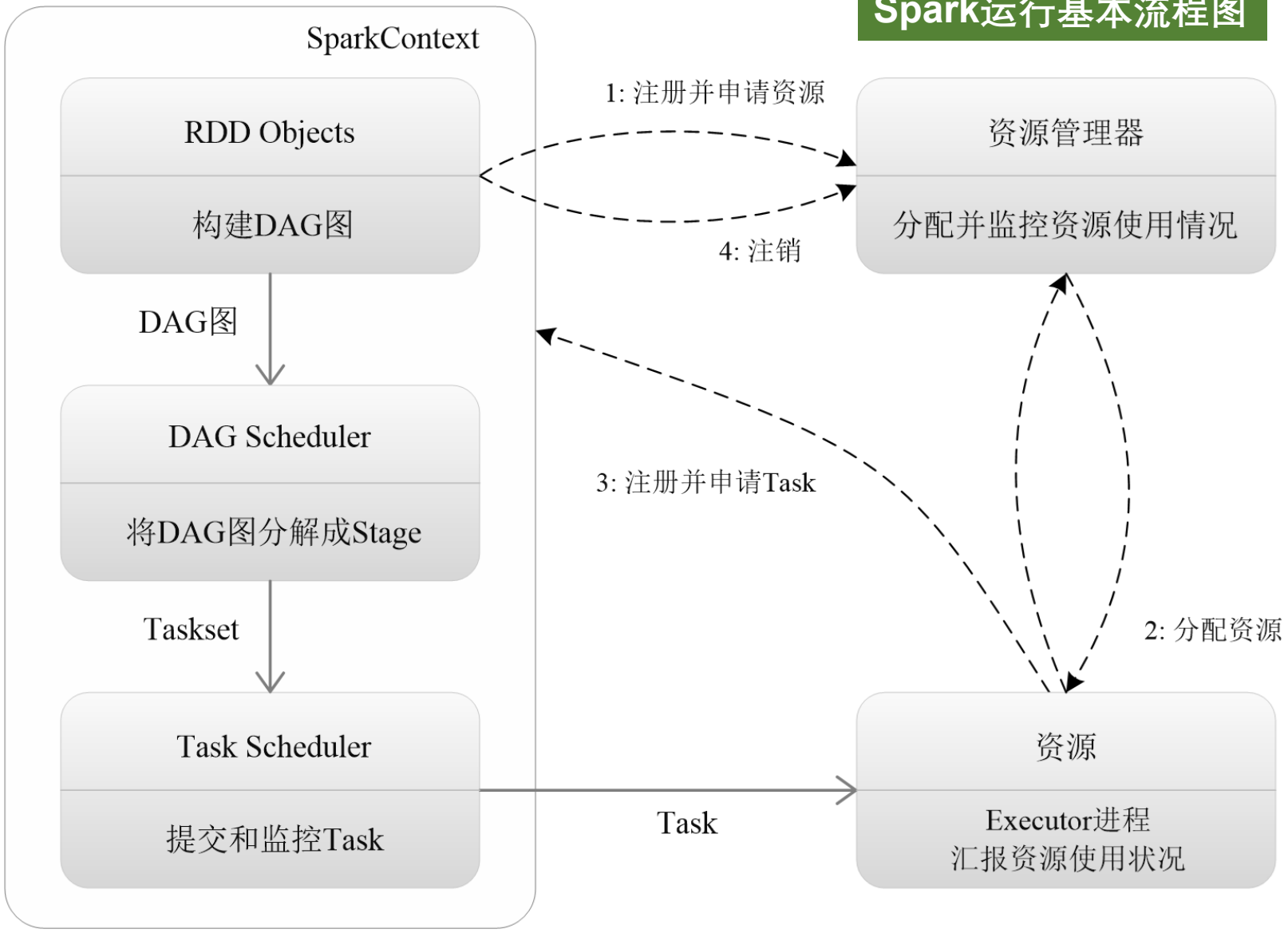


2.3.2 Spark运行基本流程

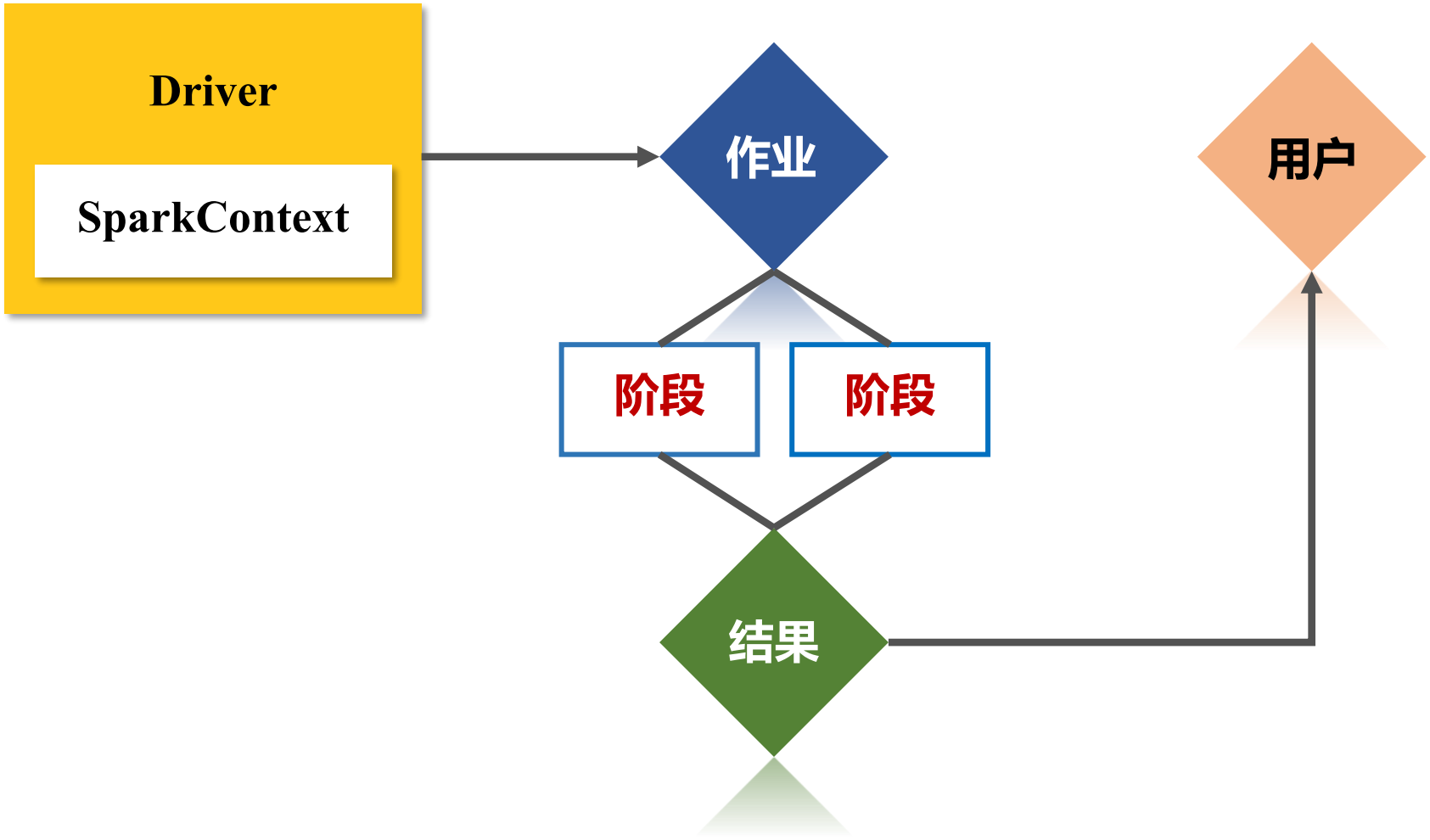
2.3.2 Spark运行基本流程



Spark运行基本流程图



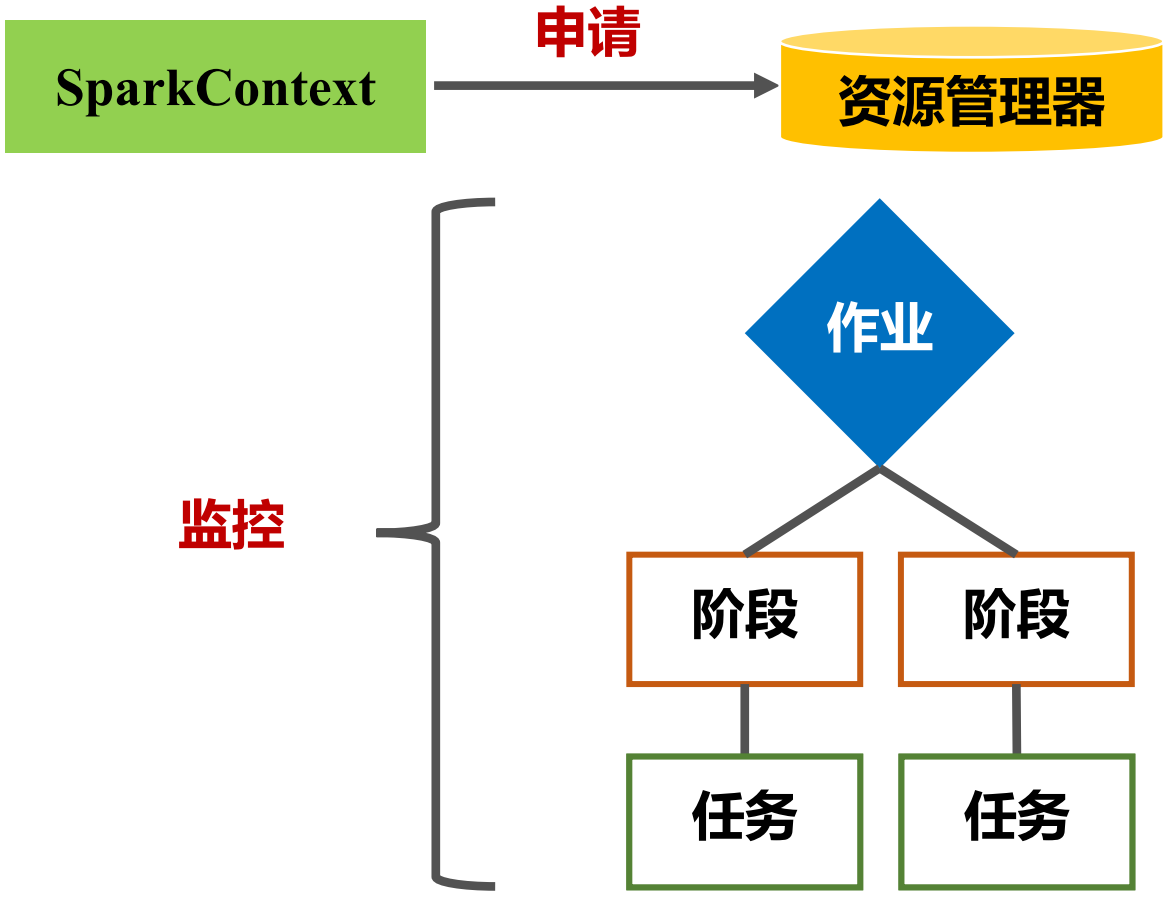
2.3.2 Spark运行基本流程

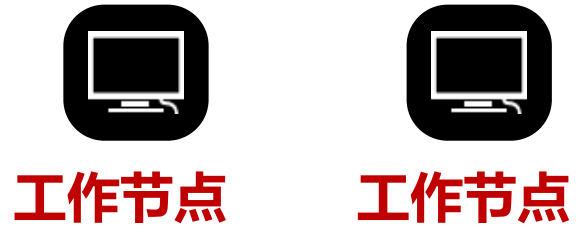


2.3.2 Spark运行基本流程



2.3.2 Spark运行基本流程

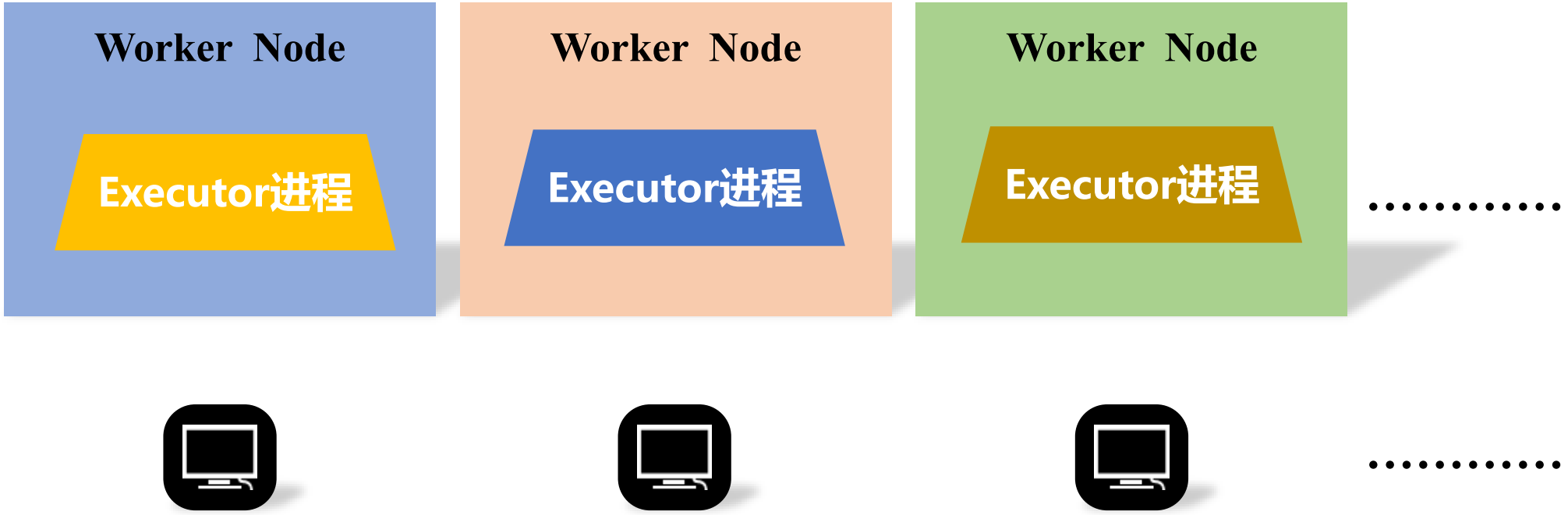




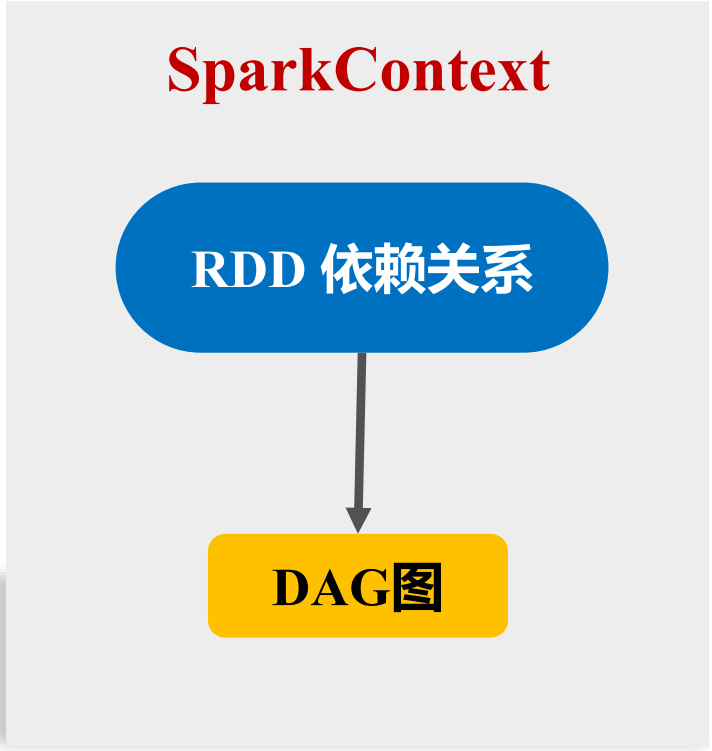
2.3.2 Spark运行基本流程

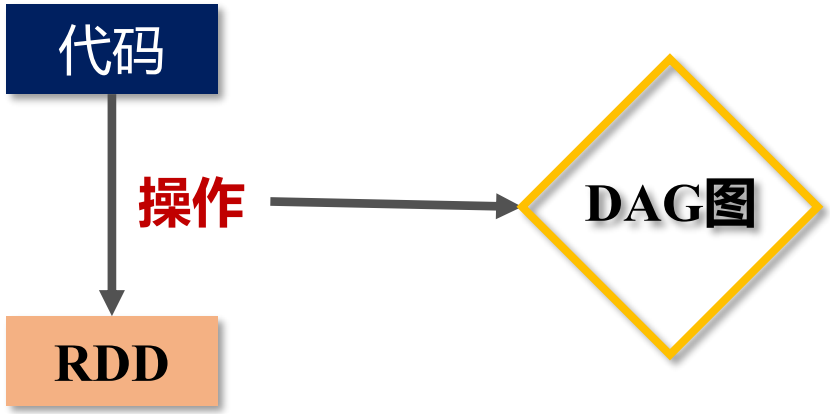


2.3.2 Spark运行基本流程

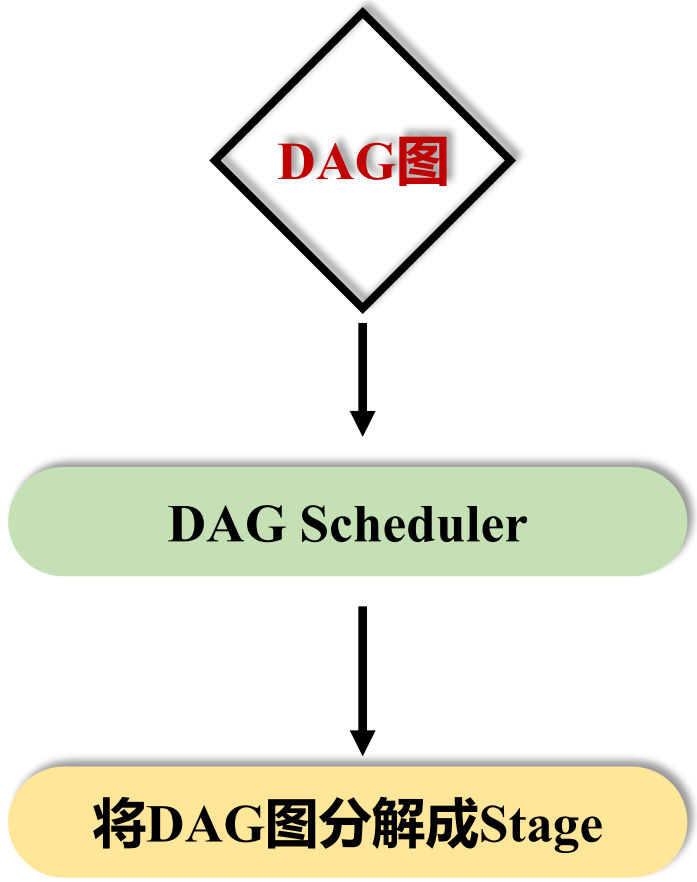


2.3.2 Spark运行基本流程

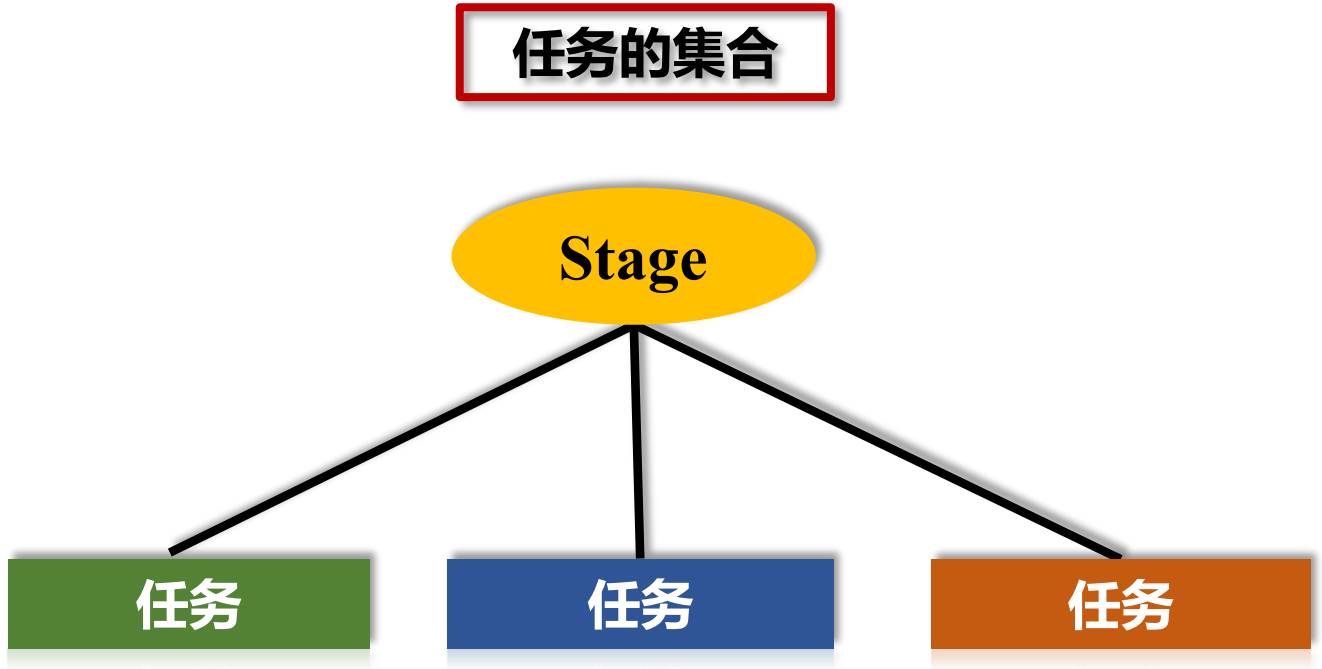




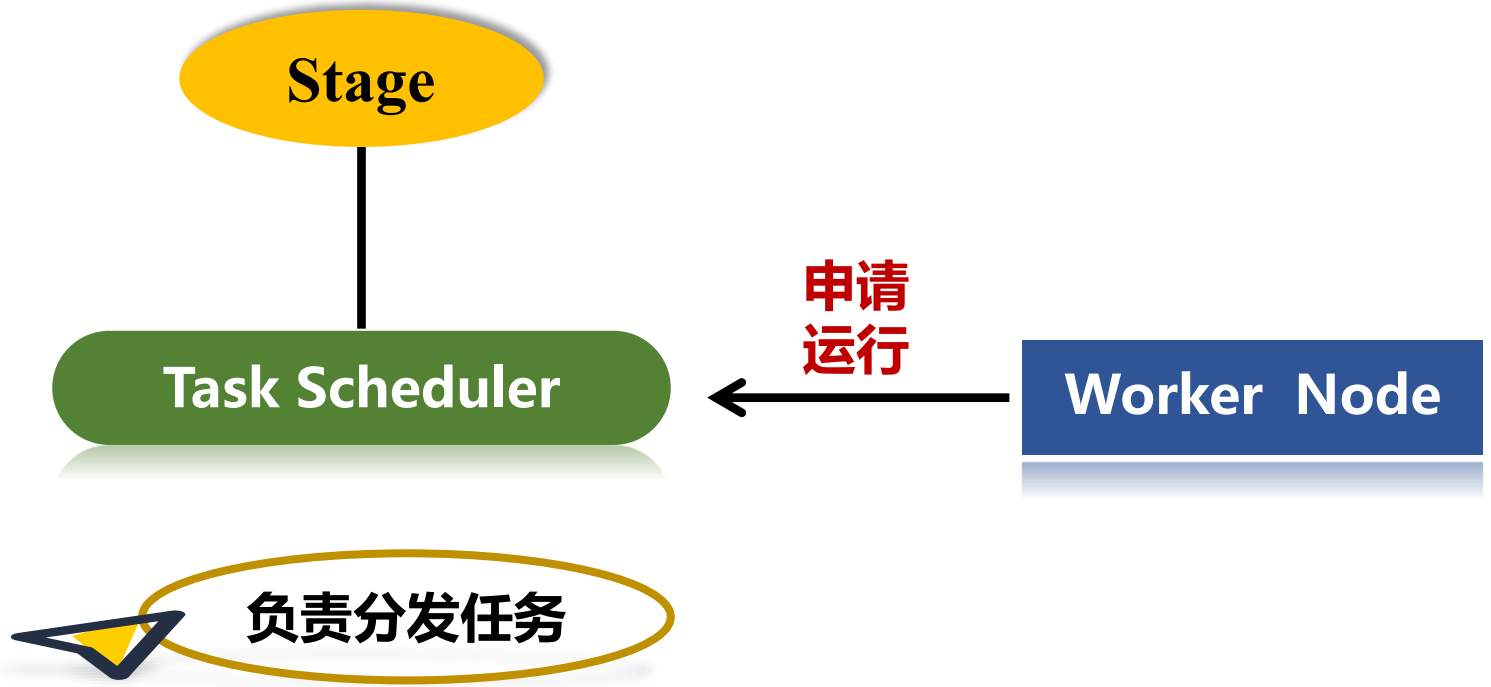
2.3.2 Spark运行基本流程



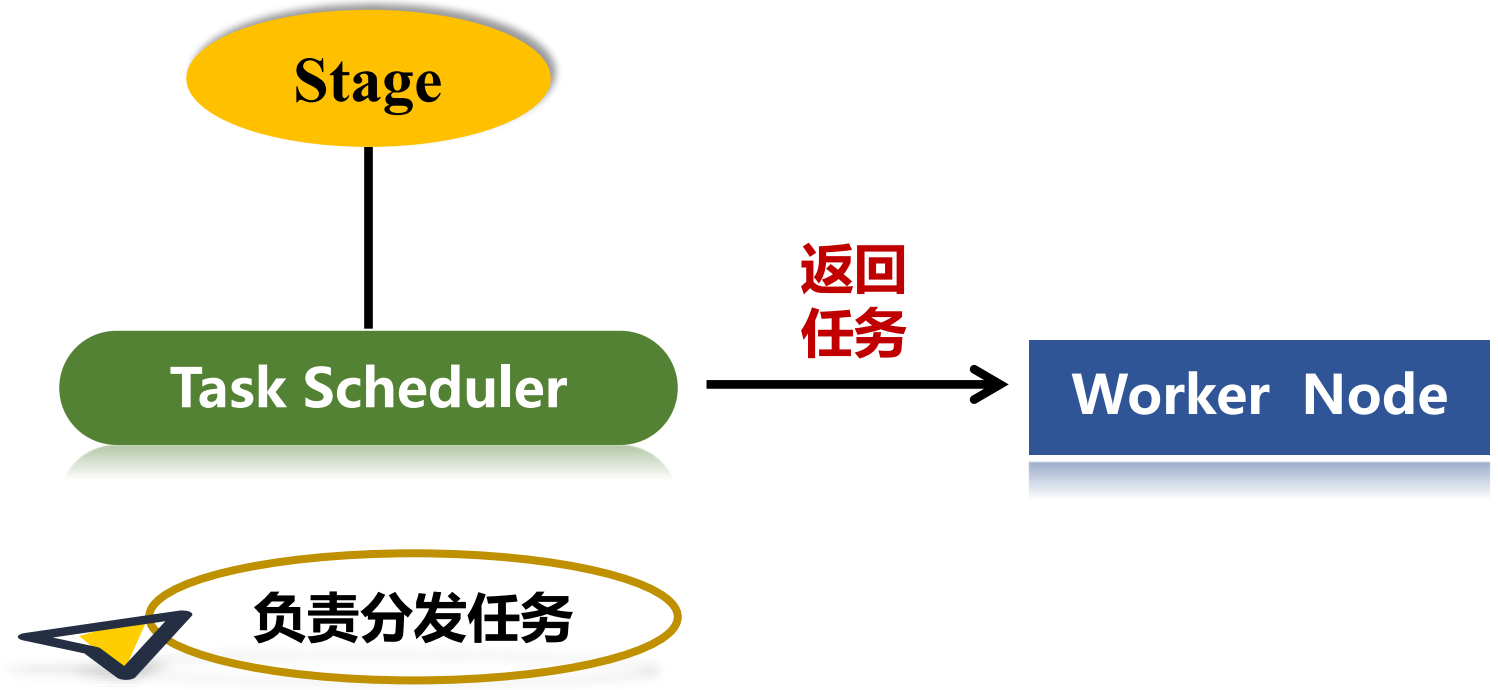
2.3.2 Spark运行基本流程



2.3.2 Spark运行基本流程



2.3.2 Spark运行基本流程

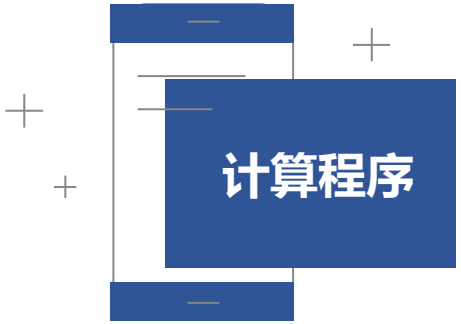


2.3.2 Spark运行基本流程



计算向数据靠拢

将数据传输到计算程序所在的机器







机器A

机器 B

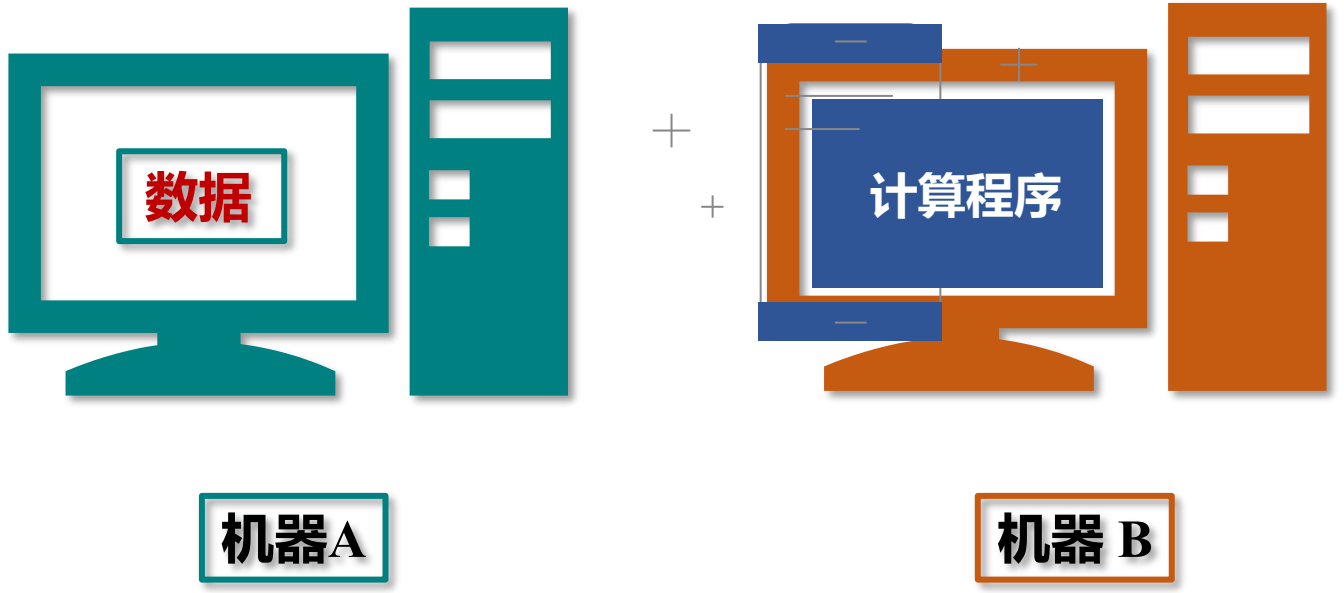


任务分发

2.3.2 Spark运行基本流程

计算向数据靠拢

将数据传输到计算程序所在的机器



任务分发

2.3.2 Spark运行基本流程

计算向数据靠拢

将数据传输到计算程序所在的机器





机器A

机器 B



任务分发

2.3.2 Spark运行基本流程

计算向数据靠拢





2.3.2 Spark运行基本流程

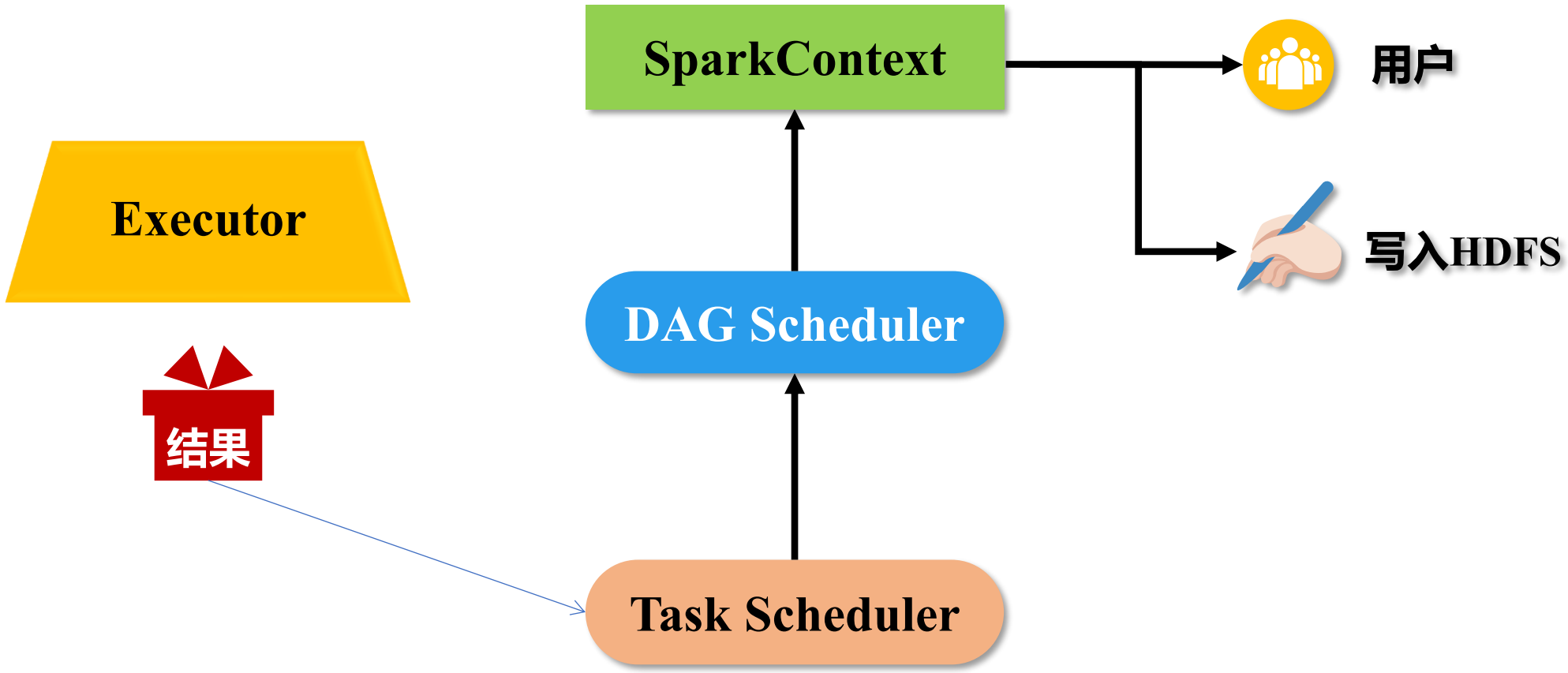
计算向数据靠拢

Task Scheduler

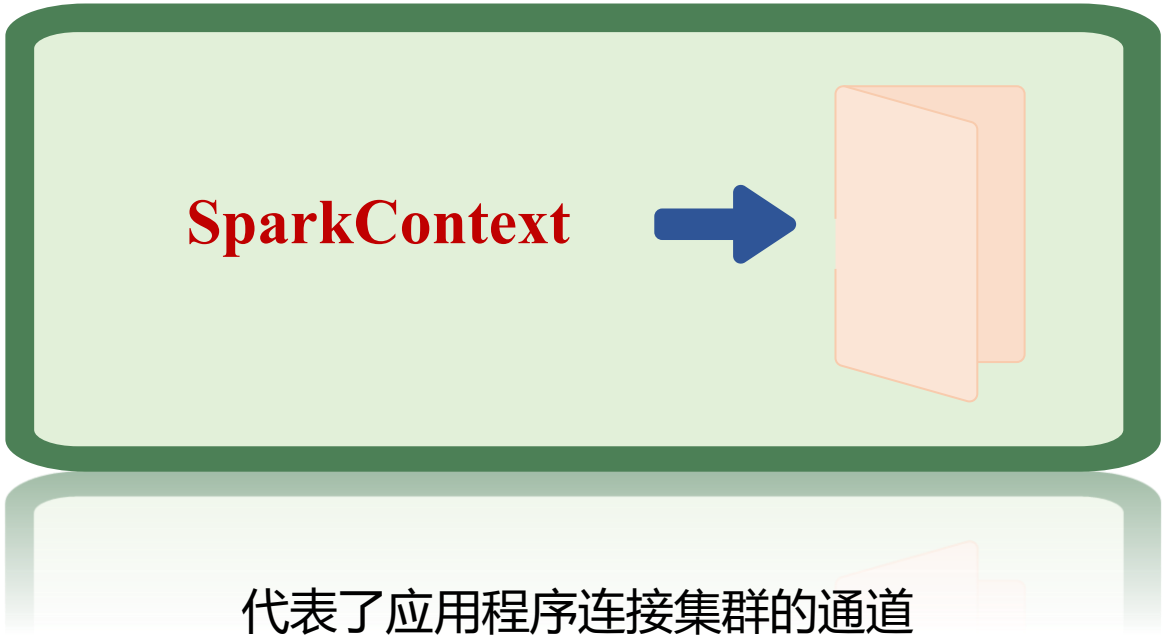




2.3.2 Spark运行基本流程

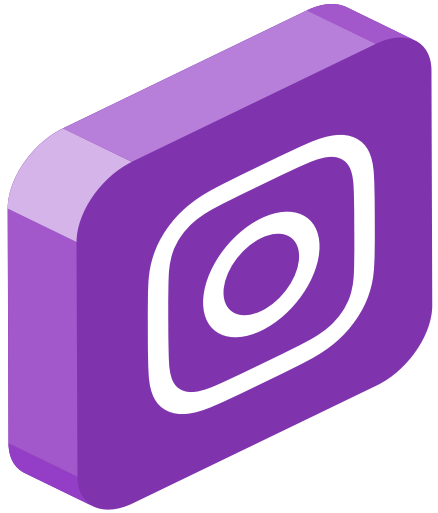


2.3.2 Spark运行基本流程



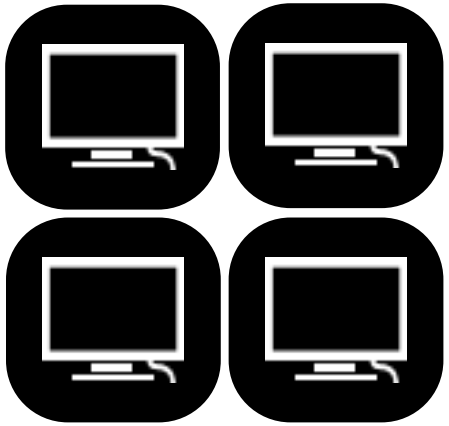
代表了应用程序连接集群的通道

2.3.2 Spark运行基本流程



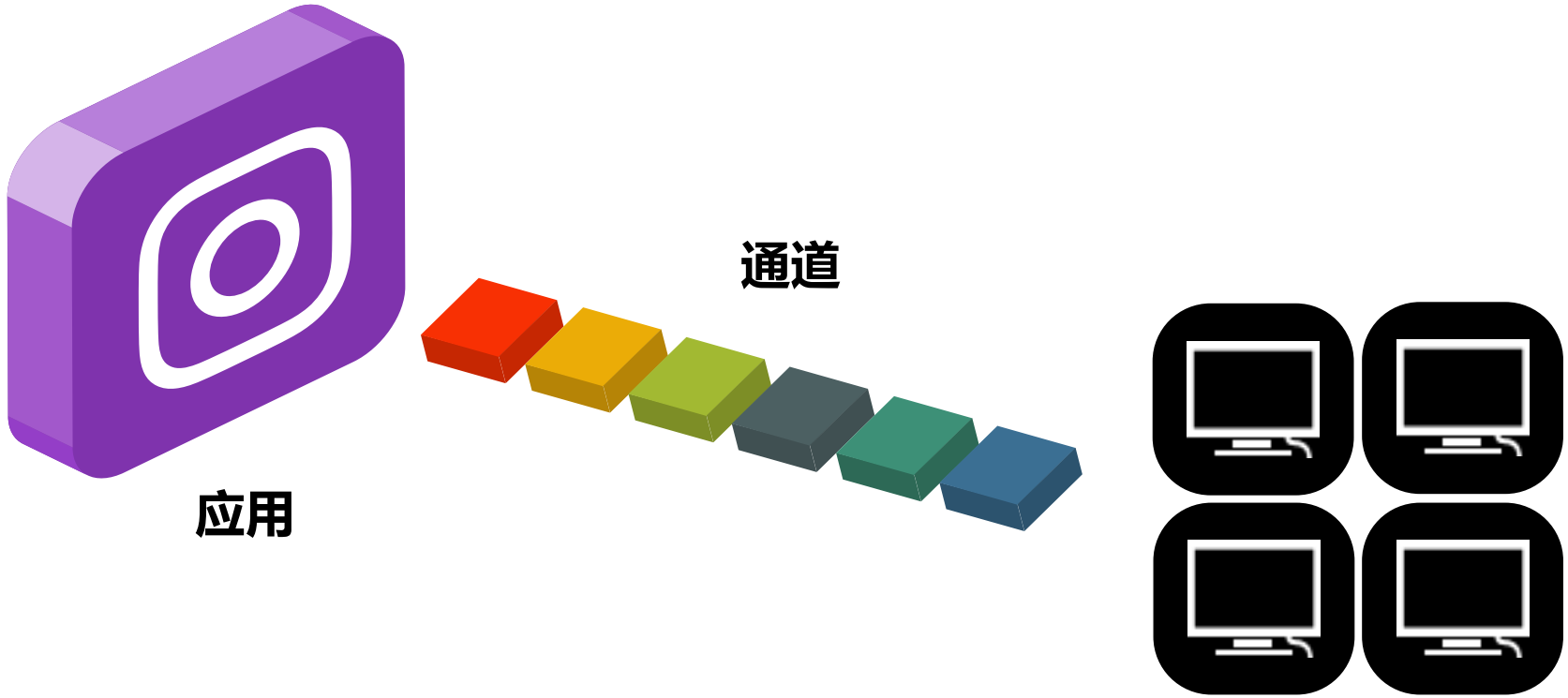
应用





集群

2.3.2 Spark运行基本流程



2.3.2 Spark运行的基本流程理解测验（选择题）



第1题

在Spark运行基本流程中，SparkContext的主要作用是什么？

A. 执行具体的计算任务
B. 代表应用程序连接集群的通道
C. 管理集群中的存储资源
D. 监控任务的执行状态

2.3.2 Spark运行的基本流程理解测验（选择题）



第2题

DAG Scheduler的主要职责是什么？

A. 将Task分发到各个Worker Node
B. 管理集群的计算资源
C. 将DAG图分解成多个Stage
D. 构建RDD之间的依赖关系

2.3.2 Spark运行的基本流程理解测验（选择题）



第3题

在Spark的运行流程中，以下哪个组件负责具体Task的分发？

A. SparkContext
B. DAG Scheduler
C. Task Scheduler
D. 资源管理器

2.3.2 Spark运行的基本流程理解测验（选择题）



第4题

Spark采用"计算向数据靠拢"策略的主要优势是什么？

A. 提高CPU利用率
B. 减少内存使用量
C. 减少网络传输开销
D. 简化编程模

2.3.2 Spark运行的基本流程理解测验（选择题）



第5题

在Spark的架构中，Executor进程主要运行在哪里？

A. Driver节点上
B. SparkContext中
C. Worker Node上
D. 资源管理器中

2.3.2 Spark运行的基本流程理解测验（选择题）



第6题

以下哪个说法关于Stage的描述是正确的？

A. Stage是单个RDD的操作
B. Stage是任务的集合，包含多个可以并行执行的Task
C. Stage只能包含一个Task
D. Stage是DAG图的另一个名称

2.3.2 Spark运行的基本流程理解测验（选择题）



第7题

在Spark运行流程中，RDD操作首先会被构建成什么？

A. Task集合
B. Stage序列
C. DAG图
D. Executor列表

2.3.2 Spark运行的基本流程理解测验（选择题）



第8题

资源管理器（如YARN、Mesos）在Spark运行中的作用是什么？

A. 构建DAG图
B. 分配集群资源并启动Executor进程
C. 执行具体的计算任务
D. 收集计算结果

2.3.2 Spark运行的基本流程理解测验（选择题）



第9题

在Spark的任务调度过程中，Task Scheduler会优先将计算任务分配到哪里？

A. CPU性能最强的节点
B. 内存最大的节点
C. 数据所在的节点
D. 负载最轻的节点

2.3.2 Spark运行的基本流程理解测验（选择题）



第10题

以下哪个是Spark运行基本流程的正确顺序？

A. SparkContext → Task Scheduler → DAG Scheduler → Executor
B. SparkContext → DAG Scheduler → Task Scheduler → Executor
C. DAG Scheduler → SparkContext → Task Scheduler → Executor
D. Task Scheduler → DAG Scheduler → SparkContext → Executor



Thank you